Landsat Science Team / October 2013

Integrating Field-Level Biophysical Metrics Derived from Landsat Science Products into a National Agricultural Data Warehouse

Jim Hipple, PhD
Physical Scientist / Remote Sensing Specialist

USDA Risk Management Agency
Office of Compliance
Strategic Data Acquisition & Analysis (SDAA)

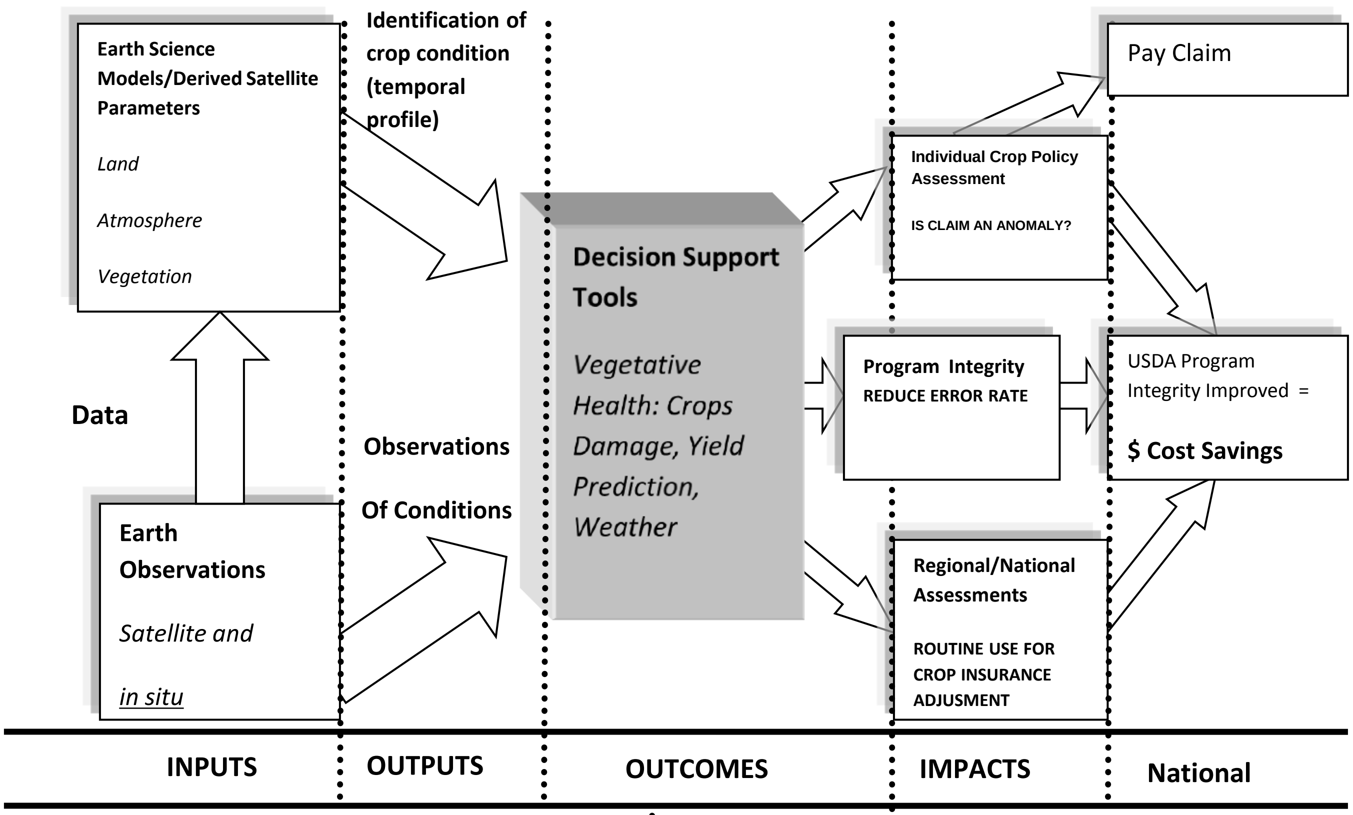
Bert Little, PhD
Professor of Computer Science & Mathematics

Associate Vice President of Academic Research and Grants
Tarleton State University, Stephenville, Texas

USDA RMA
Risk Management Agency

CAE

Applications Approach to Integrated Systems Solutions Architecture



Improved Federal Crop Insurance Program Integrity with National Impact

FY2014 Programmatic Focus

- Data integration
- Two major applications focus:
  - ‘Prevented Planting’ mapping
  - Automated claims analysis

Landsat 5, 7 & 8 Integration

- incorporate pixel-level Landsat Science Products – surface reflectance, derived biophysical metrics
- build temporal profile of key satellite derived parameters at the individual <u>field level</u> (mean, median, variability) for each image/date

Landsat Integration

- Current Work Plan
  - Preprocessing: Surface reflectance (LEDAPS) for L5, L7 & L8
    - New data on demand
    - Batch historic data
  - The output would be a 30-120 meter data consisting of:
    - Surface Reflectance bands Landsat bands (30m)
    - Surface Temperature Landsat band (60-120m)
    - Masks layers (cloud, sensor, etc.)
    - Indices
      - (NDVI, LSWI, NDWI, MSAVI, SATVI, NDTI, CRC, EVI)
  - Generating a single DOY grid for US contiguous
  - Best way of handling processed data – Teradata/Oracle Spatial
    - Currently processing MODIS pixel level as vectors in Teradata
  - Data volume

Prevented Planting Mapping

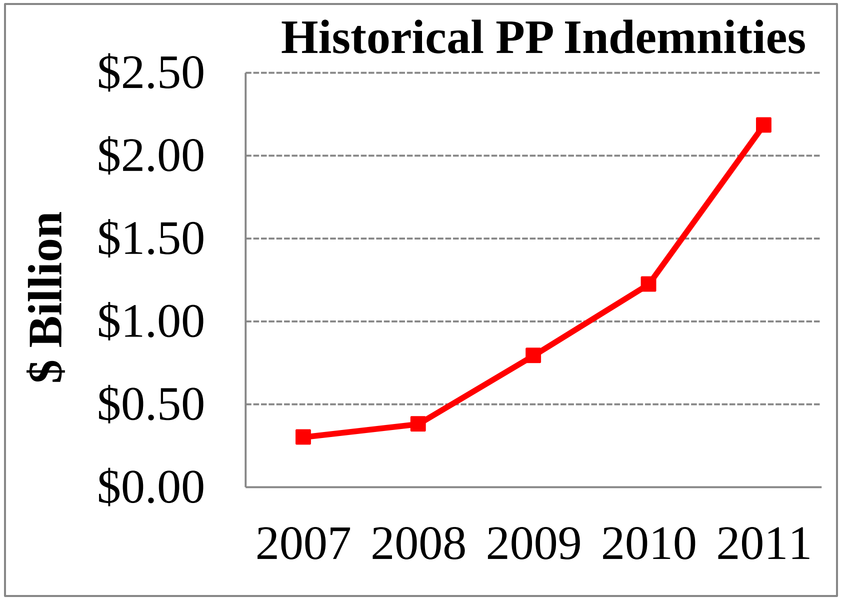
- Definition from the Basic Provisions (11-BR):
  - Failure to plant the insured crop by the final planting date designated in the Special Provisions for the insured crop in the county, or within any applicable late planting period, due to an insured cause of loss that is general to the surrounding area and that prevents other producers from planting acreage with similar characteristics. Failure to plant because of uninsured causes such as lack of proper equipment or labor to plant acreage, or use of a particular production method, is not considered prevented planting.

Prevented Planting Mapping

"Any acreage not planted to a crop that is insured under the authority of the Federal Crop Insurance Act, that is grown in the county on insurable acreage, and harvested in at least one of the four most recent crop years, using recognized good farming practices, unless such acreage was planted to an insured crop that was damaged by an insured cause of loss and adjusted for purposes of a claim under the Federal crop insurance program;"

2011 Prevented Planting Indemnities

- As of June 2012, PP Payments for 2011 = $2.19 Billion; YTD 2011 indemnities of $10.82 Billion.
  - Primary PP Cause of Loss = Excess Moisture/ Precipitation
    - $1.84 Billion = 91% of total PP payments
  - Top 3 PP States
    - North Dakota 51%
    - South Dakota 16%
    - Minnesota 4%



What does PP look like?

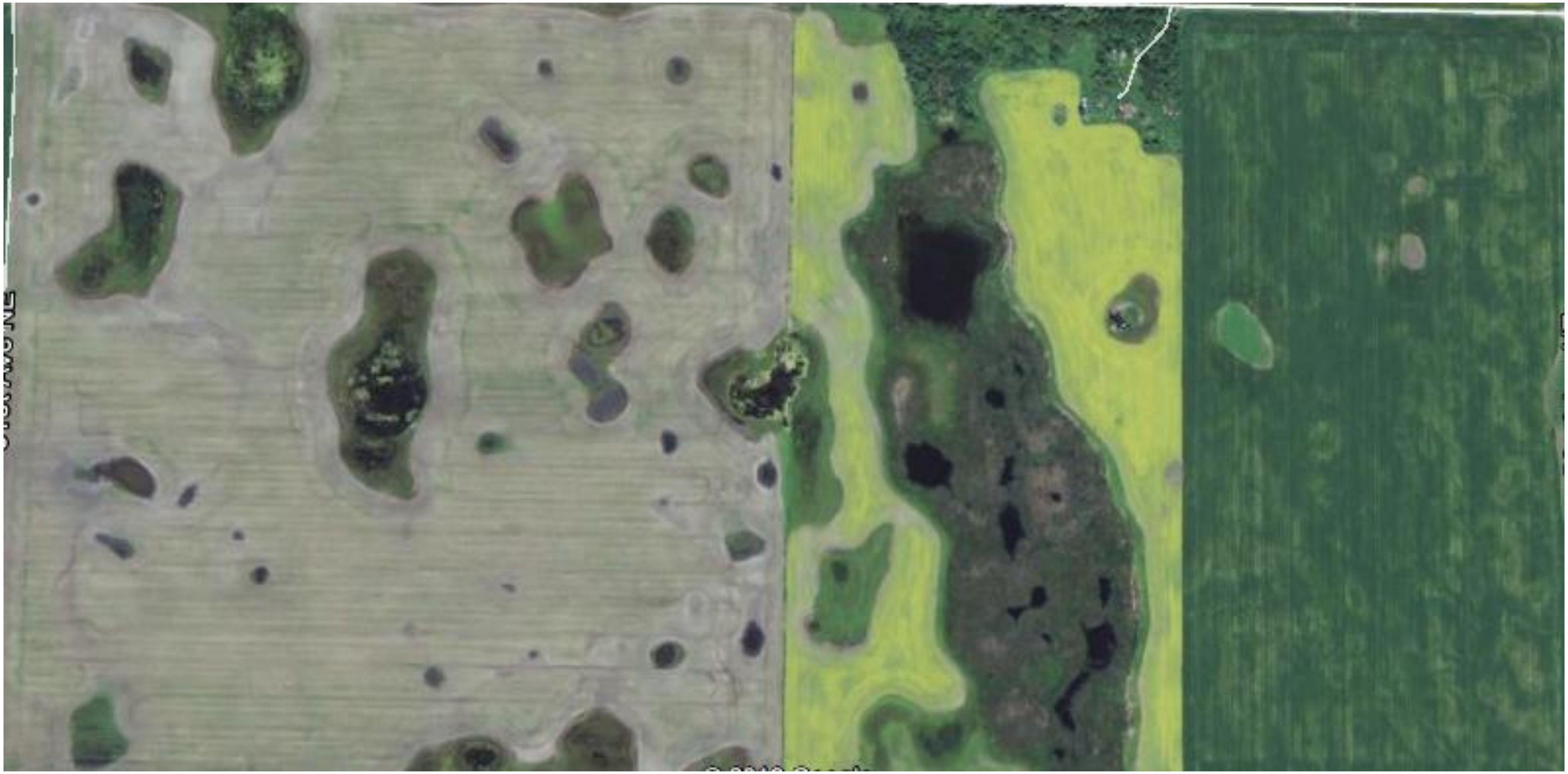


Image Source: USDA FSA NAIP

Prevented Planting Mapping

- Project is structured around the following tasks:
  – use Landsat imagery to identify standing water and water saturated soil during the planting window (April-June)
  – identify areas in standing water or water saturated soil four years in a row, three years, two years
  – determine accuracy and reliability be determined using acceptable scientific practice
  – automate the process

Automated Claims Process

- Single step process where a Compliance/RO/AIP can verify information related to a producer submitted claim
- Develop for RMA/AIP verification of claims
  - Target initially the simplified claim process
  - Process that verifies COL against distinct data type
  - Possible verifications:
    - Was acreage reported reasonably accurate (report vs. CLU acreage)?
    - Was the field prepared for planting by Producer Reported Planting date?
    - Did a crop develop?
    - Was the crop type accurate?
    - Was the Producer Reported Cause of Loss verifiable from ancillary data sets? (weather, satellite, soils)

An Example of a Claim for Verification

Crop Timeline Summary
(as reported to Insurance Company)

- Grower Reported Planting Date: December 29, 2000
- Grower Reported Acreage: 647.9 acres
- RMA Final Planting Date: February 15, 2001
- Grower Reported Cause of Loss Date: February 17 – 21, 2001

  Cause of Loss: precipitation (excess), cold-wet weather

  February 28 – March 2, 2001

  precipitation (excess), cold-wet weather

  April 7, 8, 19, 20, 2001

  hail

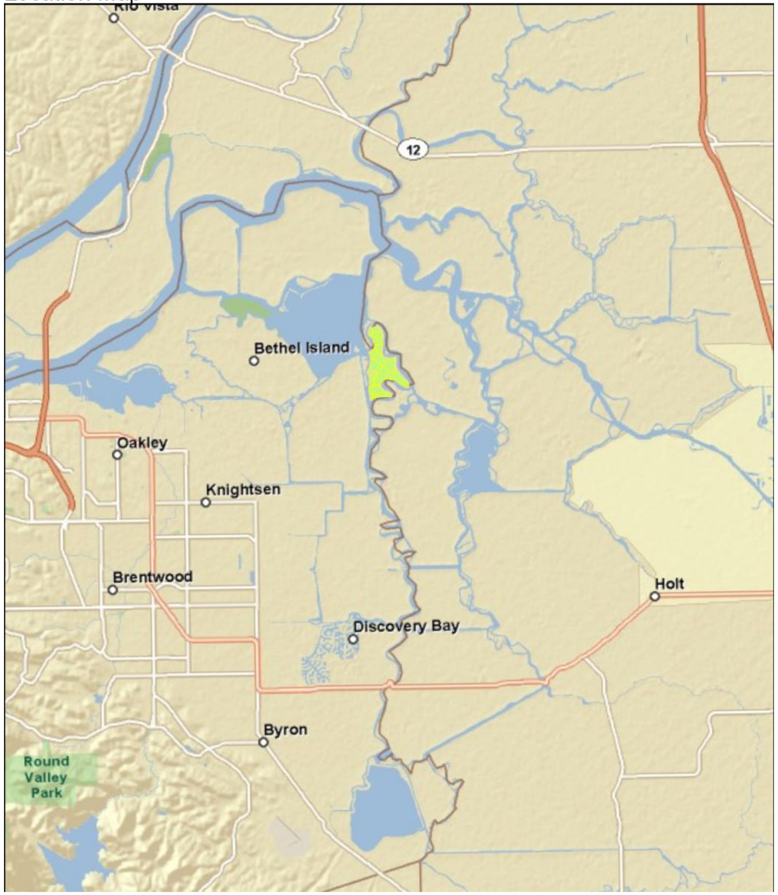
  April 16-22, 2001

  wind

  April 19-21, 2001

  precipitation (excess)
- Loss Adjustment Appraisal Date: April 23, 2001

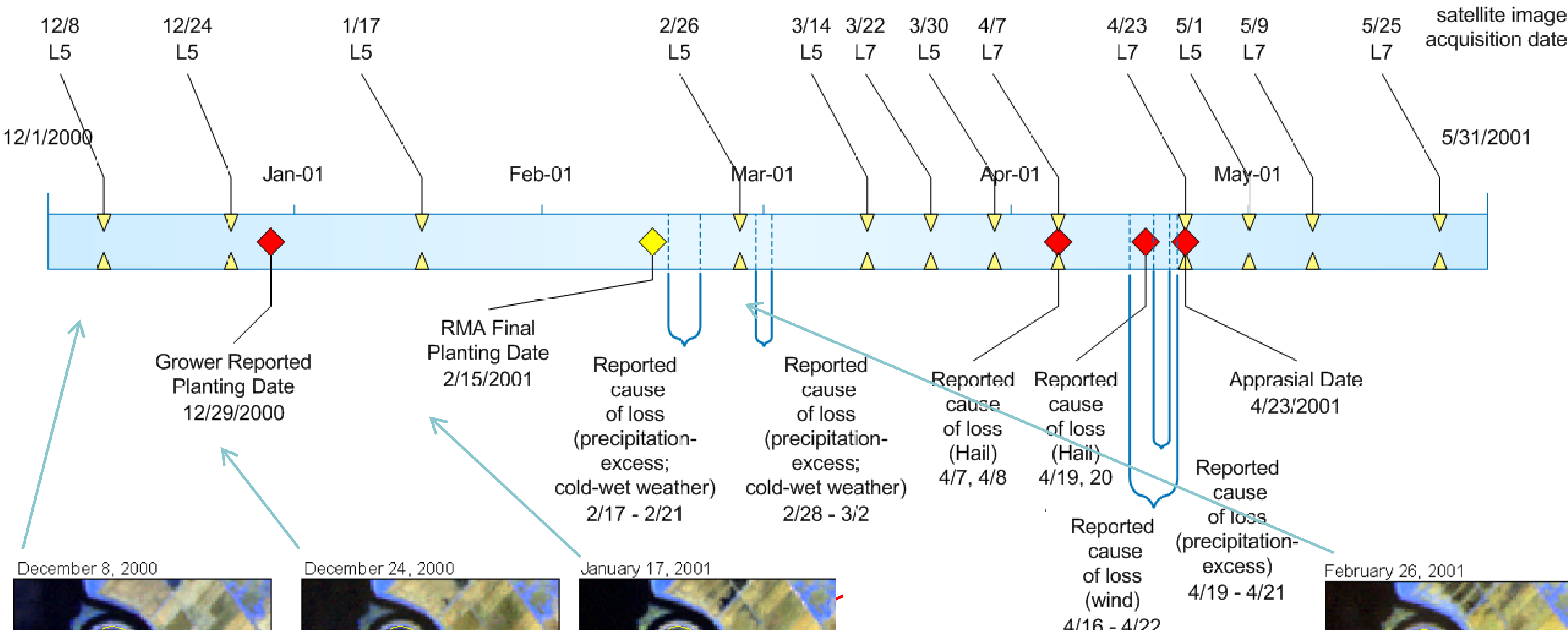
Location Map



Crop & Satellite Image Timeline

EXAMPLE

Crop Year 2001 Timeline: WHEAT

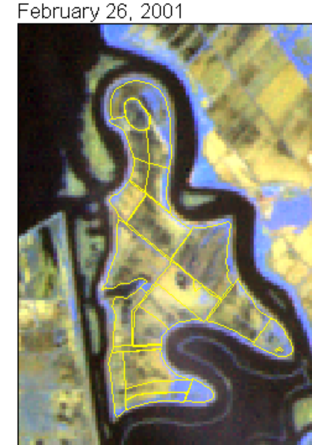


December 8, 2000

December 24, 2000

January 17, 2001

February 26, 2001

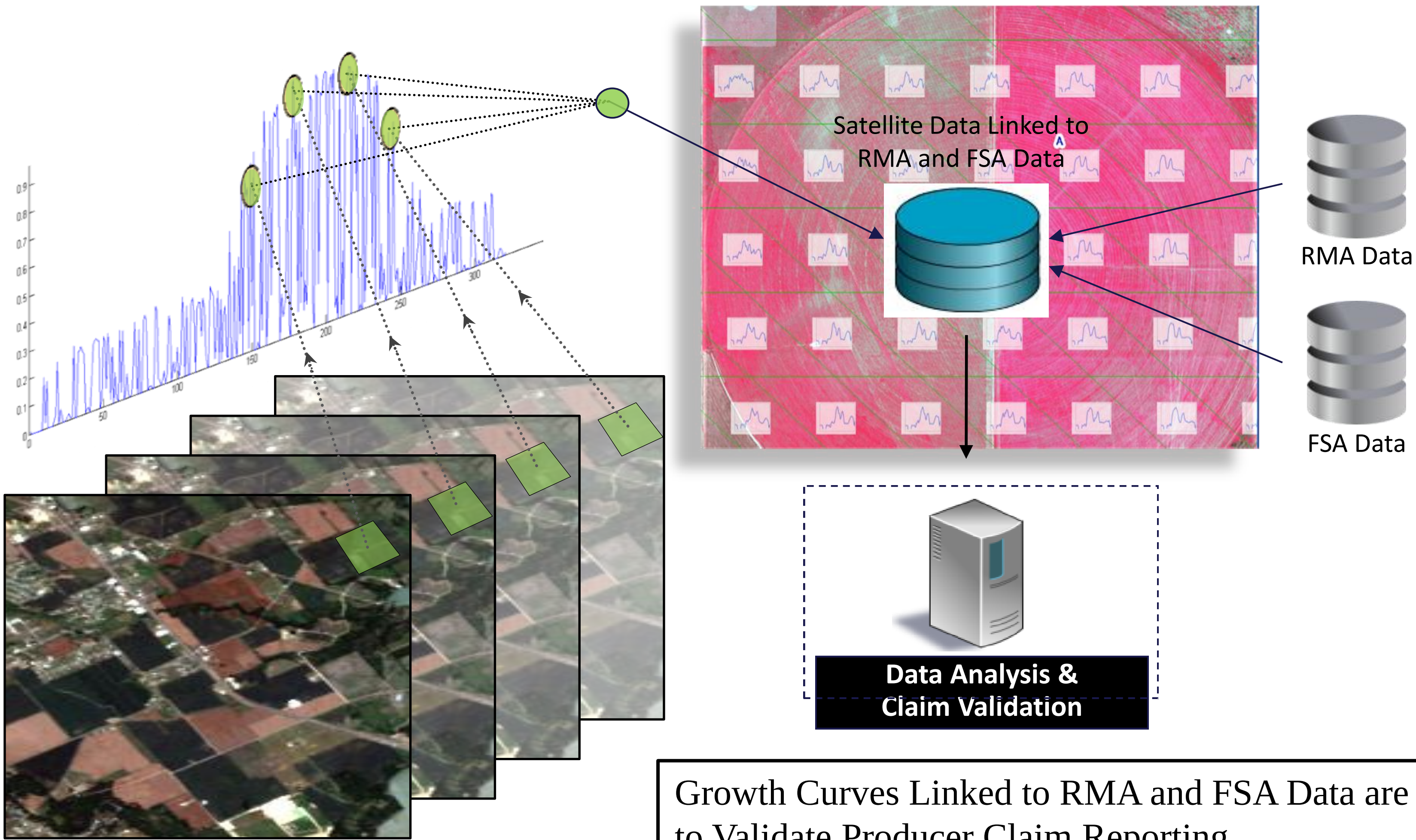


Standing water
and water saturated soil is evident on numerous fields December 8, 2000 through January 17, 2001.

The area under standing water and water saturated soil increases through January 17, 2001.

Most of the standing water or water saturated soil is gone by February 26, 2001.

Automated Claims

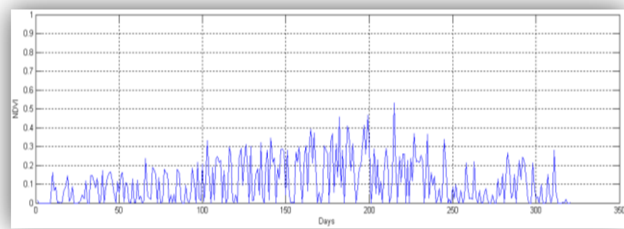


Daily MODIS Data Derives Growth Curves

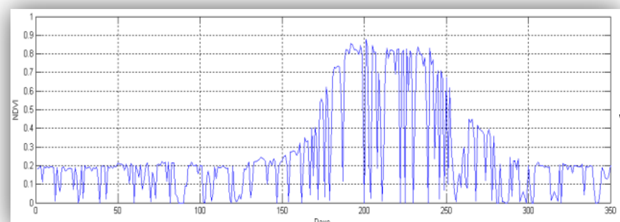
Growth Curves Linked to RMA and FSA Data are Used to Validate Producer Claim Reporting

Automated Claims Analysis

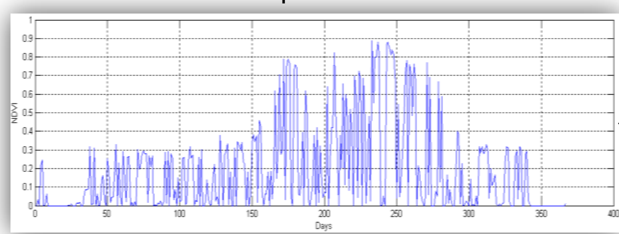
No Crop Growth

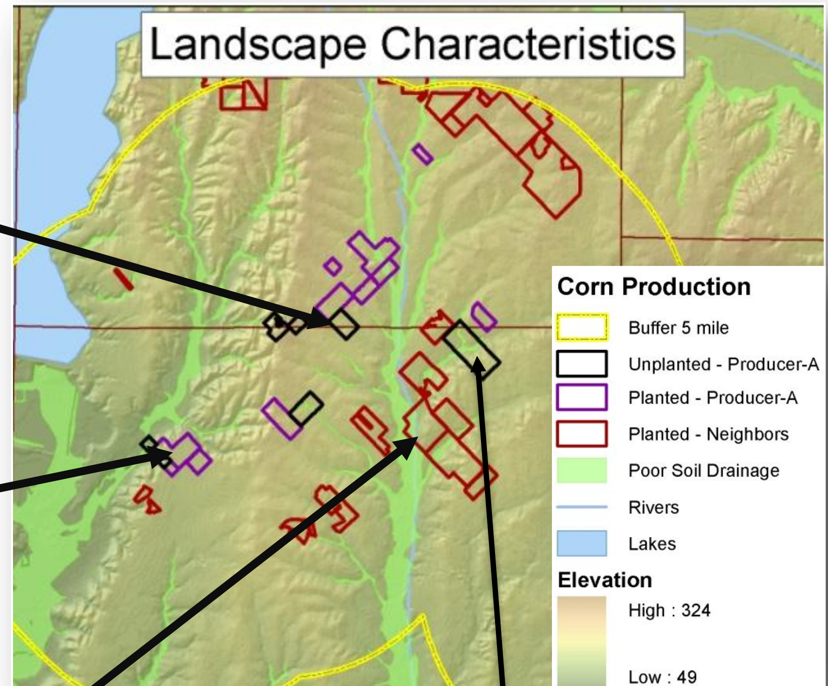


Crop Growth

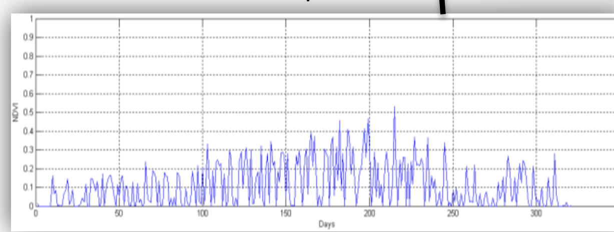


Crop Growth

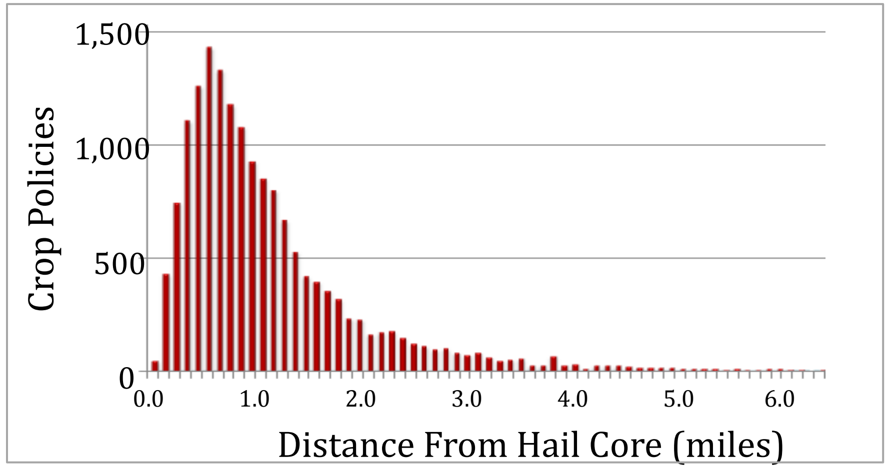




No Crop Growth



Integrate Hail Claims Validation

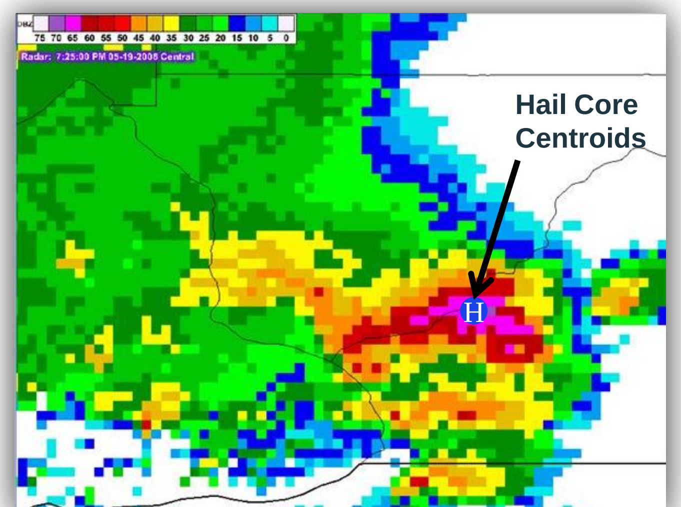


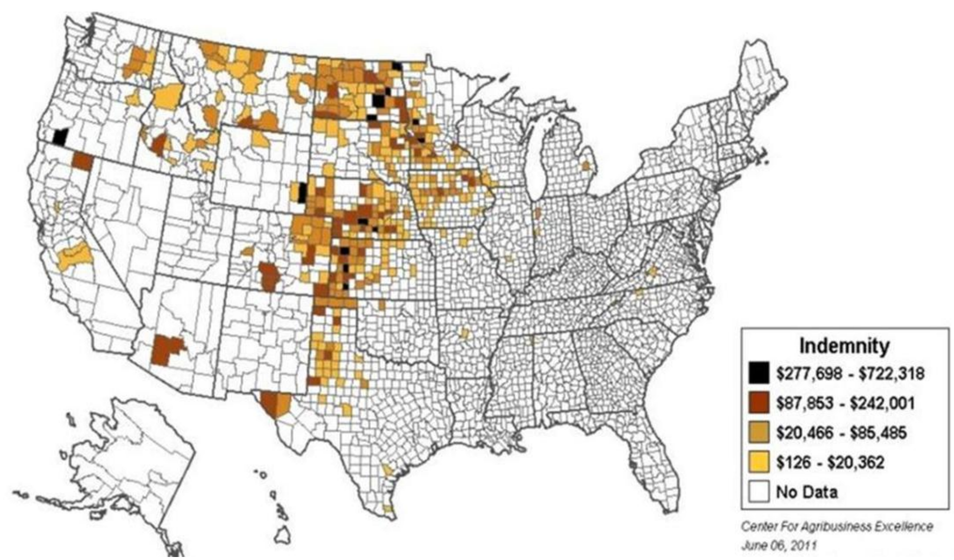
Average Hail Claims

- 0.13 Miles From a High Reflectivity Radar Value
- 1.32 Miles From the Center of the Hail Core

Anomalous Hail Claims

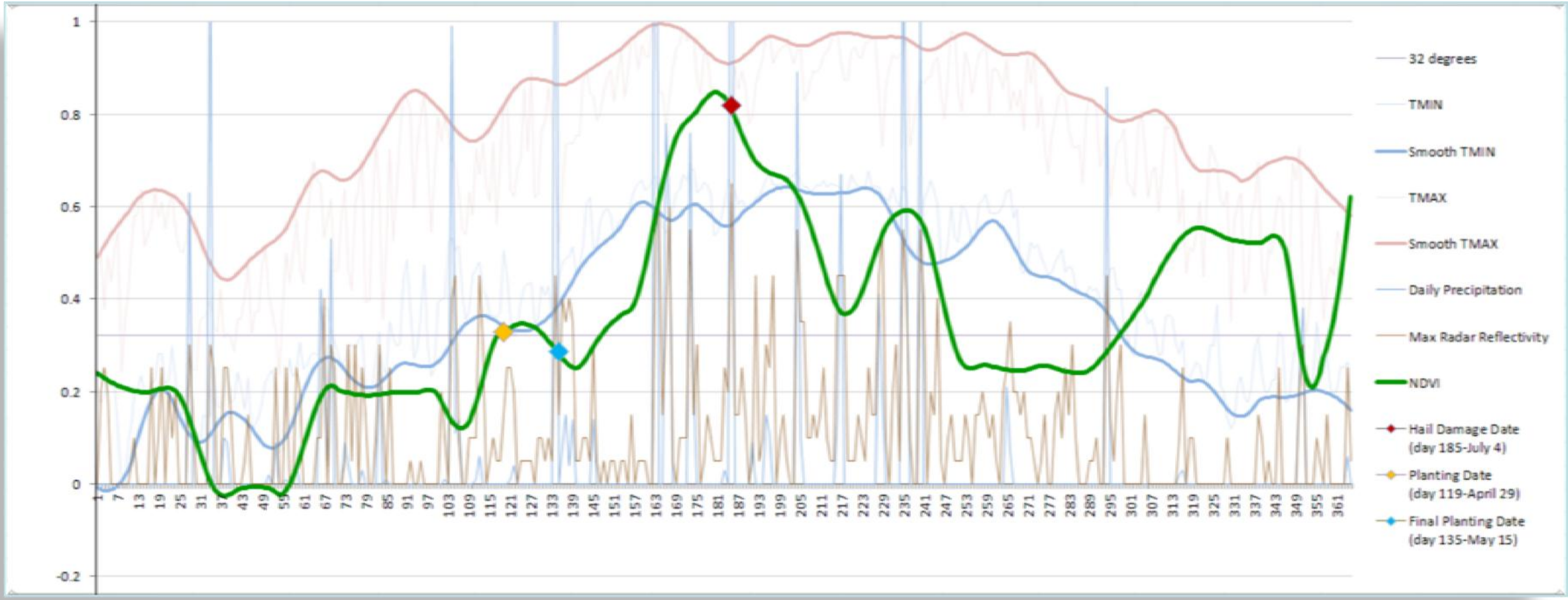
- 3.3 Miles or Greater From a Reflectivity Value Over 54
- 5.0 Miles or Greater From the Center of the Hail Core





Growth Curve & Weather Data

- Field/Pixel Level Annual Growth Curve and Weather Data Graph
- Only Available for Large Fields with Multiple MODIS Grid Cells
  - Weather Data and Growth Curve Aggregated to Field Level
    - NDVI, TMAX, TMIN, Daily Precipitation, Max Radar Reflectivity, and RMA Dates



Simplified Claims Development

(Mock Report)

- Integration of weather, satellite, soils (productivity potential) into automated reporting
  - Investigator inputs location identifier (CLU, PLSS)
  - Selects crop year
  - Report is run & PDF produced to validate claims

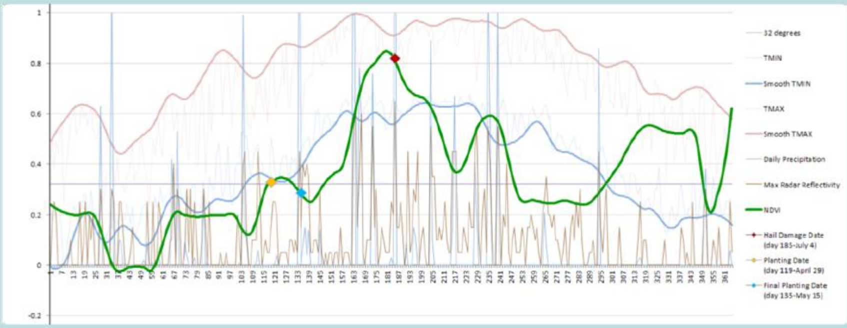
<u>Simplified Claims Loss Report</u>

Producer Name
Crop Year
FSN/Tract Field ID

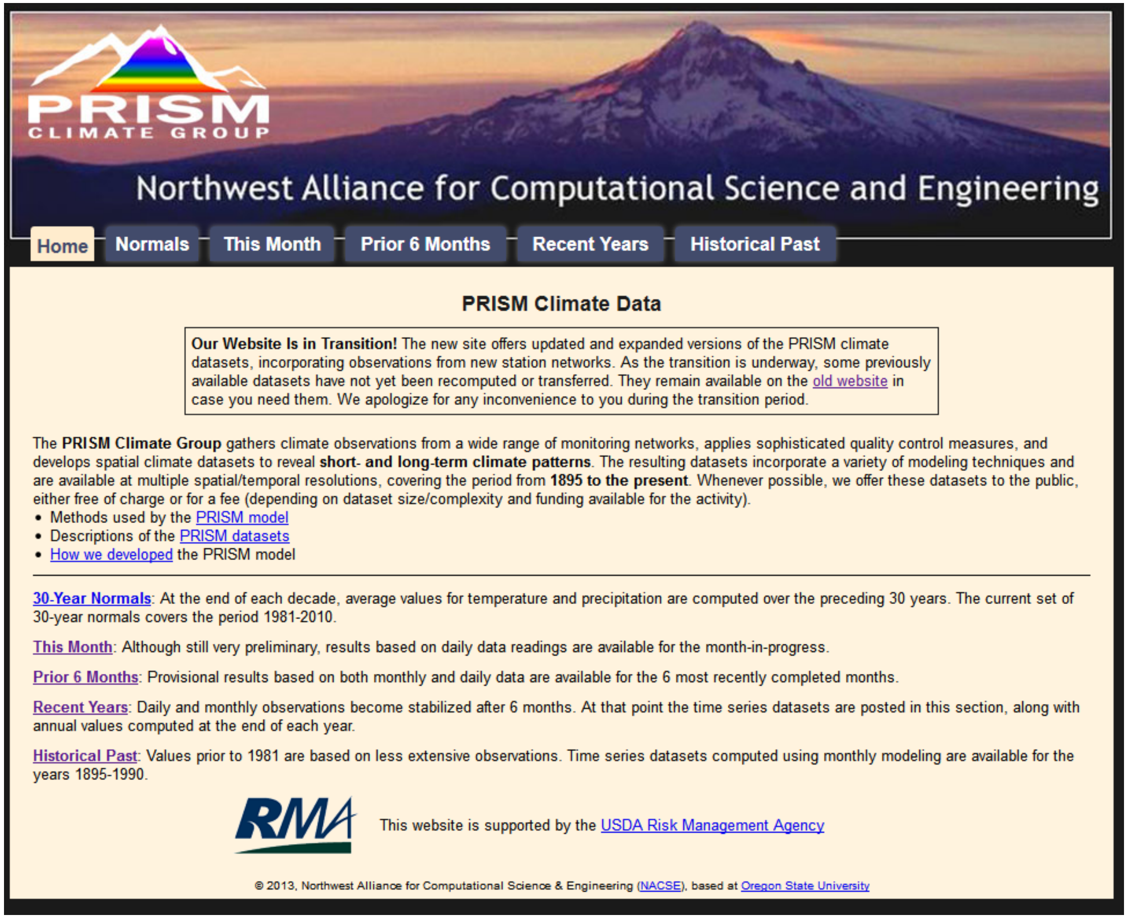
History & Reported
Claims & Yield History
Reported Causes of Loss

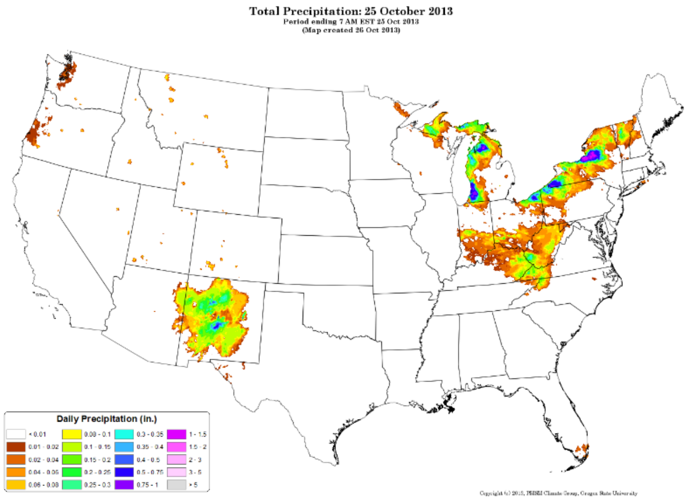
- Hail probable on these dates
- Standing water probable on these dates
- Producer may/may not have planted by RMA final planting date

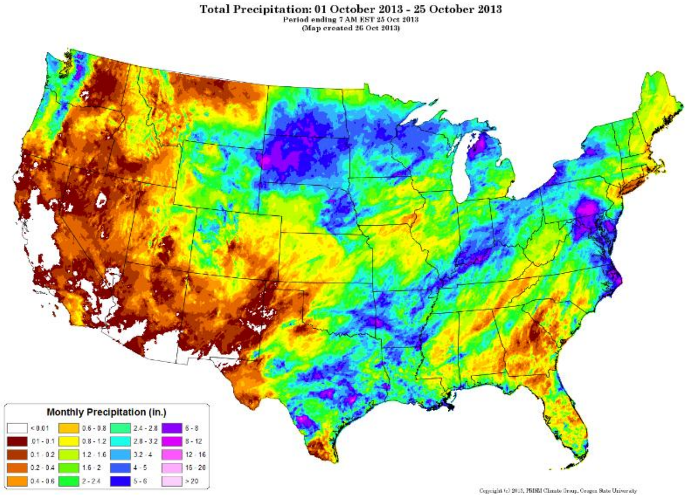
NDVI & Weather



Climate/Weather Projects with SDAA







http://prism.oregonstate.edu/

POC: Dr. Chris Daly, OSU-NACSE

Landsat 8

GCN AWARDS

PROJECT AT A GLANCE

NAME OF THE PROJECT: Crop Insurance Program Compliance and Integrity Data Warehouse

OFFICE/DIVISION/TEAM: USDA Risk Management Agency and the Center for Agribusiness Excellence at Tarleton State University

TECHNOLOGY USED: Teradata Database 14 and custom software.

TIME TO IMPLEMENT: Started in 2000.

COST: $50.68 million.

BY THE NUMBERS:

170 data sources

3 terabytes of RMA policy information

120 terabytes of weather, satellite and other remotely sensed data

1.3 million crop insurance policies

3,200 counties

"We're doing all of this in-house. When you're doing exploratory studies you have to build your own tools."

-- BERT LITTLE, TARLETON STATE UNIVERSITY

CAE RMA

Agriculture's high-res view of fraud

System combines Landsat imagery and weather data with crop insurance claims data and agricultural data to keep farmers honest

BY PATRICK MARSHALL

It doesn't have a catchy name like Batman or the Green Lantern, but the Crop Insurance Program Compliance and Integrity Data Warehouse is an effective, and innovative, crime fighter. It combs through mountains of data looking for atypical patterns among insurance claims, cross-checking them with data from high-solution satellite images and weather records. At stake are billions of dollars.

The project, run by the Agriculture Department's Risk Management Agency and developed and maintained by Tarleton State University's Center for Agribusiness Excellence, was designed to identify fraudulent crop insurance claims.

That's a more challenging task than it might seem at first glance. After all, the Federal Crop Insurance Corporation, which is overseen by RMA, has more than a million policies outstanding in 3,200 counties. When drought afflicts farms in West Texas or floods drown corn fields in Iowa, sending agents out to confirm each claim is simply not feasible.

Concerned about fraudulent crop-loss claims, Congress passed the Agriculture Risk Protection Act of 2000 (ARPA), which mandated the use of a data warehouse and data-mining technologies to improve crop insurance program compliance and integrity. Accordingly, RMA, which had already been moving in that direction, launched its data-mining project with Tarleton State's Center of Agribusiness Excellence.

The team started with the basics, collecting and comparing claims data and looking for unusual patterns. Is one farmer making claims that are different than those coming from other farms in the region?

When the program detects such a pattern, the unit will send out a letter saying that a representative from USDA may come out at some point during the year and inspect the farm's operation.

"After notifying the farmers, we saw pretty drastic behavioral changes in the producers and in their claim rates," said Kirk Bryant, deputy director of strategic data acquisition and analysis at RMA. "After we sent a letter or inspected their farms, their claims were consistent with the other claims in the county."

While the first "spot check list" was generated in 2001 solely from claims data, the program has since added data from many different sources.

The first step was to add data collected by the Farm Service Agency, including aerial imagery, crop data and information about farm loans and disaster assistance. "Through the data-mining facility we could do 'scrubbing,' and match the data between FSA and RMA," Bryant said.

The project next added data from the Natural Resources Conservation Service, which conducts soil surveys.

In 2006 the team began to integrate satellite data. At first, the data was supplied by NASA's MODIS (Moderate Resolution Imaging Spectroradiometer) satellite. "We wanted to be able to use an objective measure of vegetative health to compare against crop claims," said Bert Little, executive director of Tarleton State University's Center for Agribusiness Excellence. "In 2008 we put out a preliminary paper showing that we could tell the difference between irrigated and non-irrigated farming practices in cotton in West Texas."

With the launch of the Landsat 8 satellite earlier this year, the project has gained access to higher resolution images and data, including near-red and infrared scans. "What that gives you back is essentially the greenness that is reflected by chlorophyll in plant leaves," Little said. "The greener that signal, the healthier the plant." That can help show if there was a viable crop on land a farmer claimed he was not able to plant.

And it's not just a matter of detecting plants, since a field bordered with trees or overgrown with weeds could produce a false positive. Thanks to higher-resolution data from Landsat 8, however, the project can now distinguish between systematic growth, which is indicative of crops, and chaotic growth from weeds.

"We've written code so that the computer can go back and evaluate the satellite signal from fields," Little said.

RMA and Center for Agribusiness Excellence have built a data warehouse – which resides on Tarleton's Texas campus and runs on Teradata Database 14 – that draws data from more than 170 data sources, including 3 terabytes of RMA policy information that has been connected to 120 terabytes of weather, satellite and other remotely sensed data collected by the university. Apart from using Teradata Database 14 platform, software development has taken place at the Center for Agribusiness Excellence. "We're doing all of this in-house," Little said. " Off-the-shelf software is good for routine tasks, "but when you're doing exploratory studies you have to build your own tools."

THE PAYOFF

To date, USDA has spent $50.68 million on the program. According to RMA, the spot-check-list project alone has resulted in savings of $975 million in unjustified claims payments from 2001 through 2012. What's more, it is estimated that the program has saved $2.5 billion in cost avoidance.

While the primary payoff has been in preventing fraudulent claim payments, the system has also benefited some farmers who would incorrectly have been denied claims. In one instance, two farmers were initially denied their claims for hail damage because the National Oceanic and Atmospheric Administration could not verify that a hail storm had occurred on the day in question. The Center for Agribusiness Excellence, however, was able to locate recorded NEXRAD radar data in the data warehouse that indicated a very isolated, very heavy storm that produced the damage.

The program also has served to demonstrate the effectiveness of data mining to insurance companies. Once insurers saw the results being generated by the program, "they wanted to direct their quality control programs through data mining as opposed to doing random sampling," Bryant said. "It is so much more effective, and everything is cost-benefit driven."

"We have come light years since we started this process," said Michael Hand, RMA's deputy administrator for compliance. "Back in the beginning all we knew about remote-sensing tools was we'd see a pretty image every now and then of a farm. Now we are actually using the data from the satellites and incorporating them in our business processes."

NEXT STEPS

Officials at RMA and the Center for Agribusiness Excellence expect more benefits as the available data improves.

Bryant, in fact, sees the capabilities the team is developing being used for many other jobs in addition to preventing fraudulent crop claims. "In the future, we're looking to use this data to begin to do some proactive work in identifying problem areas in the country with different crops," he said.

And the quality of data is improving quickly. Little said his first priority is to integrate more of the Landsat 8 data. A single pixel of data from the older MODIS satellite covers roughly 11 to 13 acres, but a single pixel of data from Landsat covers a circle approximately 50 feet in diameter.

With the higher resolution data, "We can do what we're doing much better -- and we can do more specific things," he said. He also expects that the day is not far off when the program will be able to differentiate among different types of crops. "Each crop has its own special signature of reflected light," and satellite-based sensors can pick up that data, he said.

"What we're doing is bringing more and more empirical evidence into the crop insurance program so that those naysayers who claim that it's rife with waste, fraud and abuse won't have a leg to stand on," Little said. •

Tarleton CAE & RMA Awards

2013 Innovation in Big Data IT; Government Computing News, Washington DC

2013 Best Practices in Data Warehousing; TDWI (The Data Warehouse Institute), San Diego

2012 Excellence in Data Mining, Ventana Research, San Jose

2012 EPIC Data Mining Award,; Transparency, Compliance & Governance; Washington DC

2010 IEEE Computer Society, Top 10 Best Data Mining Case Study, Sydney, Australia

Questions?

James D. Hipple, PhD

USDA Risk Management Agency

Office of Compliance

Strategic Data Acquisitions & Analysis Staff

Phone: (202) 297-9328

Email: james.hipple@rma.usda.gov